

Outage Behavior of Discrete Memoryless Channels Under Channel Estimation Errors

Pablo Piantanida[†], Gerald Matz[‡], and Pierre Duhamel[†]

[†] Laboratoire des Signaux et Systèmes
CNRS/Supélec

F-91192 Gif-sur-Yvette, France

E-mail: {piantanida,pierre.duhamel}@lss.supelec.fr

[‡] Institute of Communications and Radio-Frequency Engineering
Vienna University of Technology

A-1040 Wien, Austria

E-mail: g.matz@ieee.org

Abstract

Classically, communication systems are designed assuming perfect channel state information at the receiver and/or transmitter. However, in many practical situations, only an estimate of the channel is available that differs from the true channel. We address this channel mismatch scenario by introducing the notion of estimation-induced outage capacity, for which we provide an associated coding theorem and its strong converse, assuming a discrete memoryless channel. The transmitter and receiver strive to construct codes for ensuring reliable communication with a quality of service (QoS), in terms of achieving a target rate with small error probability, no matter which degree of accuracy channel estimation arises during a transmission. We illustrate our ideas via numerical simulations for transmissions over Ricean fading channels using rate-limited feedback channel and maximum likelihood (ML) channel estimation. Our results provide intuitive insights on the impact of the channel estimate and the channel characteristics (SNR, Ricean K-factor, training sequence length, feedback rate, etc.) on the mean outage capacity.

1. Introduction

Channel uncertainty, caused e.g. by time variations/fading, interference, or channel estimation errors, can severely impair the performance of wireless systems. Even if the channel is quasi-static and interference is small, uncertainty induced by imperfect channel state information (CSI) remains. This motivates us to study the design of communication systems which require to ensure information transmission at a target rate satisfying a quality of service (QoS), i.e. reliable communication, no matter which degree of accuracy channel estimation arises during the communication.

We first review the model for communication under channel uncertainty over a discrete memoryless channel

(DMC) with finite input alphabet \mathcal{X} and output alphabet \mathcal{Y} [1]. A specific instance of the unknown channel is characterized by a transition probability mass (PM) $W(\cdot|x, \theta) \in \mathcal{W}_\Theta$ with an unknown channel state $\theta \in \Theta \subseteq \mathbb{C}^d$. Here, $\mathcal{W}_\Theta = \{W(\cdot|x, \theta) : x \in \mathcal{X}, \theta \in \Theta\}$ is a family of conditional transition PMs on \mathcal{Y} , parameterized by a vector $\theta \in \Theta$. By considering a composite channel model, through the notion of reliable communication based on the average of the error probability over all channel estimation errors. Capacity bounds for additive white Gaussian noise (AWGN) channels with MMSE channel estimation, i.e. imperfect CSI at the receiver (CSIR), and no CSI at the transmitter (CSIT) were derived in [2].

Throughout the paper we assume that the channel state, which neither the transmitter nor the receiver knows exactly, is constant within blocks of duration T symbol periods (coherence time), and channel states in different blocks are i.i.d. $\theta \sim \psi(\theta)$. The extension of the DMC $W(\cdot|x, \theta)$ to n channel uses within a block is given by $W^n(\mathbf{y}|\mathbf{x}, \theta) = \prod_{i=1}^n W(y_i|x_i, \theta)$ where $\mathbf{x} = (x_1, \dots, x_n)$ and $\mathbf{y} = (y_1, \dots, y_n)$. The receiver only knows an estimate $\hat{\theta}_R$ of the channel state and a characterization of the estimator performance in terms of the conditional probability density function (pdf) $\psi(\theta|\hat{\theta}_R)$ (this can be obtained using \mathcal{W}_Θ , the estimator function and the a priori distribution of θ). Moreover, a feedback channel provides the transmitter with noisy CSI $\hat{\theta}_T$ ($\hat{\theta}_T$ in general is different from $\hat{\theta}_R$, e.g. due to quantization). The joint distribution of $(\hat{\theta}_T, \hat{\theta}_R, \theta)$ is given by $\psi(\hat{\theta}_T, \hat{\theta}_R, \theta)$. The scenario underlying these assumptions is motivated by current wireless systems, where for the case of a mobile receiver T may be too short to permit reliable estimation of the fading coefficients.

The concept of outage capacity was first proposed in [3] for fading channels. It is defined as the maximum rate that can be supported with probability $1 - \gamma$, where γ is a prescribed outage probability. In contrast, ergodic capacity is the maximum information rate for which error probability decays exponentially with the code length.

In our setting, a transceiver using $\hat{\theta} = (\hat{\theta}_R, \hat{\theta}_T)$ instead

This work was supported by the European Union IST network of excellence NEWCOM.

of θ obviously might not support an information rate R even if R is less than the channel's capacity under perfect CSIR (even arbitrarily small rates might not be supported if $\hat{\theta}$ and θ happen to be strongly different). Consequently, outages induced by channel estimation errors will occur with a certain probability γ . The outage probability depends on the codeword error probability, averaged over a random coding ensemble and over all channel realizations given the estimated state. We first formalize the notion of estimation-induced outage capacity for general DMCs, and then we present a coding theorem providing the explicit expression for the corresponding capacity, which is a function of the outage probability γ (Section 2). Due to the independence of different blocks (coherence intervals), it is sufficient to study the estimation-induced outage rate $C(\gamma, \psi_{\theta|\hat{\theta}}, \hat{\theta})$ for a single block (coherence interval), for which the channel state is fixed but unknown to the transmitter and the receiver. Since this rate still depends on the random channel estimates $\hat{\theta}$, we will consider the performance measure

$$\bar{C}(\gamma, \psi_{\theta|\hat{\theta}}) = E_{\hat{\theta}}\{C(\gamma, \psi_{\theta|\hat{\theta}}, \hat{\theta})\}, \quad (1)$$

which describes (average) information rate with prescribed outage probability. The expectation in (1) is with respect to the joint distribution $\psi(\hat{\theta}) = \psi(\hat{\theta}_T|\hat{\theta}_R) \int_{\Theta} \psi(\hat{\theta}_R|\theta)\psi(\theta)d\theta$ and reflects an average over a large number of blocks (coherence intervals), cf. the discussion in [4].

Our notion of reliably communication is relevant e.g. for communication systems where a quality of service (QoS) in terms of error performance must be ensured although significant channel variations occur due to user mobility. An example of such a scenario involving a fading Ricean channel with AWGN, rate-limited feedback, and maximum likelihood (ML) channel estimation, will be considered in Section 4 to illustrate the mean outage capacity $\bar{C}(\gamma, \psi_{\theta|\hat{\theta}})$.

2. Problem Statement and Main Result

In this section, we first develop a proper formalization of the notion of estimation-induced outage capacity and state our main result.

2.1. Problem Definition

A message m from the set $\mathcal{M} = \{1, \dots, \lfloor \exp(nR) \rfloor\}$ is transmitted using a length- n block code defined as a pair (φ, ϕ) of mappings, where $\varphi : \mathcal{M} \times \Theta \mapsto \mathcal{X}^n$ is the encoder (that utilizes $\hat{\theta}_T$), and $\phi : \mathcal{Y}^n \times \Theta \mapsto \mathcal{M} \cup \{0\}$ is the decoder (that utilizes $\hat{\theta}_R$). The random rate, which depends on the unknown channel realization θ through its probability of error, is given by $n^{-1} \log M_{\theta, \hat{\theta}}$. The maximum (over all messages) error probability

$$e_{\max}(\varphi, \phi, \hat{\theta}; \theta) = \max_{m \in \mathcal{M}} \sum_{\mathbf{y} \in \mathcal{Y}^n: \phi(\mathbf{y}, \hat{\theta}_R) \neq m} W^n(\mathbf{y}|\varphi(m, \hat{\theta}_T), \theta).$$

For a given channel estimate $\hat{\theta} = (\hat{\theta}_R, \hat{\theta}_T)$, and $0 < \epsilon, \gamma < 1$, an outage rate $R \geq 0$ is (ϵ, γ) -achievable on an unknown channel $W(\cdot|x, \theta) \in \mathcal{W}_{\Theta}$, if for every $\delta > 0$ and every sufficiently large n there exists a sequence of length- n block codes such that the rate satisfies

$$\Pr\left(\{\theta \in \Lambda_{\epsilon} : n^{-1} \log M_{\theta, \hat{\theta}} \geq R - \delta\} \mid \hat{\theta}\right) \geq 1 - \gamma,$$

where $\Lambda_{\epsilon} = \{\theta \in \Theta : e_{\max}(\varphi, \phi, \hat{\theta}; \theta) \leq \epsilon\}$ is the set of all channel states allowing for reliable decoding. This definition requires that maximum error probabilities larger than ϵ occur with probability less than γ , i.e., $P_{\theta|\hat{\theta}}(\Lambda_{\epsilon}|\hat{\theta}) \geq 1 - \gamma$. The practical advantage of such definition is that for any degree of accuracy channel estimation, the transmitter and receiver strive to construct codes for ensuring reliable communication with probability $1 - \gamma$, no matter which unknown state θ arises during the transmission.

A rate $R \geq 0$ is γ -achievable if it is (ϵ, γ) -achievable for every $0 < \epsilon < 1$. Let $C_{\epsilon}(\gamma, \psi_{\theta|\hat{\theta}}, \hat{\theta})$ be the largest (ϵ, γ) -achievable rate for an outage probability γ and a given estimated $\hat{\theta}$. The *estimation-induced outage capacity* of this channel is then defined as the largest γ -achievable rate, i.e.,

$$C(\gamma, \psi_{\theta|\hat{\theta}}, \hat{\theta}) = \lim_{\epsilon \downarrow 0} C_{\epsilon}(\gamma, \psi_{\theta|\hat{\theta}}, \hat{\theta}).$$

2.2. Coding Theorem

We next state a coding theorem quantifying the estimation-induced outage capacity $C(\gamma, \psi_{\theta|\hat{\theta}}, \hat{\theta})$ for our scenario where an estimate $\hat{\theta}_R$ of the channel state is known at the decoder and a noisy version $\hat{\theta}_T$ of $\hat{\theta}_R$ is known at the encoder. We impose an input constraint that depends on the transmitter CSI and requires that $\Gamma(P) = \sum_{x \in \mathcal{X}} \Gamma(x)P(x)$ is less than $\mathcal{P}(\hat{\theta}_T)$. Here, $\Gamma(\cdot)$ is any arbitrary non-negative function, and $P(\cdot)$ denotes the input distribution.

Theorem 2.1 *Given $0 \leq \gamma < 1$ the estimation-induced outage capacity of an unknown DMC $W(\cdot|x, \theta) \in \mathcal{W}_{\Theta}$ is given by*

$$C(\gamma, \psi_{\theta|\hat{\theta}}, \hat{\theta}) = \max_{P: \Gamma(P) \leq \mathcal{P}(\hat{\theta}_T)} \mathcal{C}(\gamma, \psi_{\theta|\hat{\theta}}, \hat{\theta}, P), \quad (2)$$

where

$$\mathcal{C}(\gamma, \psi_{\theta|\hat{\theta}}, \hat{\theta}, P) = \sup_{\Lambda \subset \Theta: \Pr(\Lambda|\hat{\theta}) \geq 1 - \gamma} \inf_{\theta \in \Lambda} I(P, W(\cdot|\cdot, \theta)). \quad (3)$$

In addition, $C_{\epsilon}(\gamma, \psi_{\theta|\hat{\theta}}, \hat{\theta}) = C(\gamma, \psi_{\theta|\hat{\theta}}, \hat{\theta}) \quad \forall 0 < \epsilon < 1$.

In this theorem, we used the mutual information

$$I(P, W(\cdot|\cdot, \theta)) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P(x) W(y|x, \theta) \log \frac{W(y|x, \theta)}{Q(y|\theta)}$$

with $Q(y|\theta) = \sum_{x \in \mathcal{X}} P(x) W(y|x, \theta)$. We emphasize that the supremum in (3) is taken over all subsets Λ of Θ that have (conditional) probability at least $1 - \gamma$. Furthermore, codes achieving capacity (3) can be viewed as codes for a simultaneous channel \mathcal{W}_{Λ^*} , which has been determined by

the decoder. Hence, this outage capacity $C(\gamma, \psi_{\theta|\hat{\theta}}, \hat{\theta})$ is seen to equal the maximum capacity of all compound channels that are contained in \mathcal{W}_Θ and, conditioned on $\hat{\theta}$, have sufficiently high probability. The significance of Theorem 2.1 is that it provides an explicit way to evaluate the outage capacity for an unknown but estimated channel for arbitrary estimation accuracies without additional assumptions.

Observe that if perfect CSIR is available then $\Lambda_\epsilon = \Theta$ and the instantaneous mutual information is attainable. Thus, every rate R can be associated to the set $\Lambda_R = \{\theta \in \Theta : I(P, W(\cdot|\cdot, \theta)) \geq R - \delta\}$ whose probability is $1 - \gamma$. Therefore, in this case, the channel can be modeled as a compound channel, whose transition probability depends on a random parameter $\theta \in \Theta$. In the following section we provide a proof of Theorem 2.1.

3. Proof of the Coding Theorem

In this section we determine the capacity by using the tools of information theory, according to the definition in Section 2. The proof of Theorem 2.1 is based on an extension of the maximal code lemma [5] to bound the minimum size of the images for the considered channels, according to the notion of estimation-induced outage capacity.

Throughout this section, we will use the notion of (conditional) information-typical (I-typical) sets defined in terms of (Kullback-Leibler) divergence, i.e., $T_P^n(\delta) = \{\mathbf{x} : \mathcal{D}(\hat{P}_n \| P) \leq \delta\}$ and $T_{W|P}^n(\mathbf{x}, \delta) = \{\mathbf{y} : \mathcal{D}(\hat{W}_n \| W | \hat{P}_n) \leq \delta\}$ where \hat{P}_n is the empirical PM associated with \mathbf{x} and \hat{W}_n is the empirical conditional PM associated with \mathbf{x} and \mathbf{y} .

3.1. Generalized Maximal Code Lemma

Let \mathcal{J}_Λ denote the set of all common η -images $\mathcal{B}^n \subseteq \mathcal{Y}^n$ associated to a set $\mathcal{A}^n \subset \mathcal{X}^n$ via the collection of simultaneous DMCs \mathcal{W}_Λ ,

$$\mathcal{J}_\Lambda(\mathcal{A}^n, \eta) = \left\{ \mathcal{B}^n : \inf_{\theta \in \Lambda} W^n(\mathcal{B}^n | \mathbf{x}, \theta) \geq \eta \text{ for all } \mathbf{x} \in \mathcal{A}^n \right\}.$$

In the following, we will denote by

$$g_\Lambda(\mathcal{A}^n, \eta) = \min_{\mathcal{B}^n \in \mathcal{J}_\Lambda(\mathcal{A}^n, \eta)} \|\mathcal{B}^n\| \quad (4)$$

the minimum of the cardinalities of all common η -images \mathcal{B}^n . For a given channel estimate $\hat{\theta} = (\hat{\theta}_T, \hat{\theta}_R)$ with degraded CSIT $\theta \oplus \hat{\theta}_R \oplus \hat{\theta}_T$, a code $(\mathbf{x}_1(\hat{\theta}_T), \dots, \mathbf{x}_M(\hat{\theta}_T); \mathcal{D}_1^n(\hat{\theta}), \dots, \mathcal{D}_M^n(\hat{\theta}))$ according to the above definition consists of a set of codewords $\mathbf{x}_m(\hat{\theta}_T)$ and associated decoding sets $\mathcal{D}_m^n(\hat{\theta})$ (i.e., the decoder reads $\phi(\mathbf{y}, \hat{\theta}) = m$ iff $\mathbf{y} \in \mathcal{D}_m^n(\hat{\theta})$). For any set \mathcal{A}^n , we call a code admissible if $\mathbf{x}_m(\hat{\theta}_T) \in \mathcal{A}^n$, all decoding sets $\mathcal{D}_m^n(\hat{\theta}) \subseteq \mathcal{Y}^n$ are mutually disjoint, and the set

$$\Lambda_\epsilon = \left\{ \theta \in \Theta : \max_{m \in \mathcal{M}} W^n((\mathcal{D}_m^n(\hat{\theta}))^c | \mathbf{x}_m(\hat{\theta}_T), \theta) \leq \epsilon \right\}, \quad (5)$$

satisfies that $\Pr(\Lambda_\epsilon | \hat{\theta}) \geq 1 - \gamma$. Any input distribution satisfying the input constraint $\mathcal{P}(\hat{\theta}_T)$ is denoted by $P(\cdot | \hat{\theta}_T)$.

Theorem 3.1 *Let two arbitrary numbers $0 < \epsilon, \delta < 1$ be given. There exists a positive integer n_0 such that for all $n \geq n_0$ the following two statements hold.*

1) *Direct Part: For any $\mathcal{A}^n \subset T_{P|\hat{\theta}_T}^n(\delta, \hat{\theta}_T)$ and any random set $\Lambda \subset \Theta$ with $\Pr(\Lambda | \hat{\theta}) \geq 1 - \gamma$, there exists an admissible sequence of length- n block codes of size*

$$M_{\theta, \hat{\theta}} \geq \exp \left[-n(H(\mathcal{W}_\Lambda | P) - \delta) \right] g_\Lambda(\mathcal{A}^n, \epsilon - \delta), \quad (6)$$

for all $\theta \in \Lambda$, where $\Lambda_\epsilon = \Lambda$.

2) *Converse Part: For $\mathcal{A}^n = T_{P|\hat{\theta}_T}^n(\delta, \hat{\theta}_T)$, the size of any admissible sequence of length- n block codes is bounded*

$$M_{\theta, \hat{\theta}} \leq \exp \left[-n(H(\mathcal{W}_{\Lambda_\epsilon} | P) + \delta) \right] g_{\Lambda_\epsilon}(\mathcal{A}^n, \epsilon + \delta), \quad (7)$$

for all $\theta \in \Lambda_\epsilon$.

The proof of this theorem easily follows from basic properties of I-typical sequences and the concept of robust I-typical sets in Appendix A. Whereas, Theorem 2.1 is obtained through the following corollary.

Corollary 3.2 *For a given channel estimate $\hat{\theta}$ and an outage probability γ , and $0 < \epsilon, \delta < 1$ and any PM $P(\cdot | \hat{\theta}_T) \in \mathcal{P}(\mathcal{X})$. Let $\mathcal{C}(\gamma, \psi_{\theta|\hat{\theta}}, \hat{\theta}, P)$ be defined by expression (3). Then the following statements holds:*

(i) *There exists an optimal sequence of block codes of length n and size $M_{\theta, \hat{\theta}}$, whose maximum error probabilities larger than ϵ occur with probability less than γ , such that*

$$\Pr \left(n^{-1} \log M_{\theta, \hat{\theta}} \geq R - 2\delta | \hat{\theta} \right) \geq 1 - \gamma \quad (8)$$

for all rate $R \leq \mathcal{C}(\gamma, \psi_{\theta|\hat{\theta}}, \hat{\theta}, P)$, provided that $n \geq n_0$.

(ii) *For any block codes of length n , size $M_{\theta, \hat{\theta}}$ and code-words in $T_{P|\hat{\theta}_T}^n(\delta, \hat{\theta})$, whose maximum error probabilities larger than ϵ occur with probability less than γ . The largest code size satisfies*

$$\Pr \left(n^{-1} \log M_{\theta, \hat{\theta}} > R + 2\delta | \hat{\theta} \right) < \gamma \quad (9)$$

for all rate $R \geq \mathcal{C}(\gamma, \psi_{\theta|\hat{\theta}}, \hat{\theta}, P)$, whenever $n \geq n_0$.

Proof: From the direct part of Theorem 3.1 and Lemma A.1, we have that there exists admissible codes such that

$$n^{-1} \log M_{\theta, \hat{\theta}} \geq n^{-1} \log g_\Lambda(\mathcal{A}^n, \epsilon - \delta) - H(\mathcal{W}_\Lambda | P) - \delta, \quad (10)$$

for all $\theta \in \Lambda$ and sets $\Lambda \subset \Theta$ (having probability at least $1 - \gamma$). Let $\hat{\mathcal{D}}^n$ be the common $(\epsilon - \delta)$ -image of minimal size $\|\hat{\mathcal{D}}^n\| = g_\Lambda(\mathcal{A}^n, \epsilon - \delta)$. Then it is easy to show that $\inf_{\theta \in \Lambda} W_\theta P^n(\hat{\mathcal{D}}^n) \geq (\epsilon - \delta)^2$. By applying Corollary 1.2.14

in [5] to this relation and substituting it in (10), we obtain for all $n \geq n'_0(|\mathcal{X}|, |\mathcal{Y}|, \epsilon, \delta)$,

$$n^{-1} \log M_{\theta, \hat{\theta}} \geq \inf_{\theta \in \Lambda} I(P, W(\cdot, \theta)) - 2\delta, \quad (11)$$

for all $\theta \in \Lambda$, where the last inequality follows from the concavity of the entropy function with respect to W_θ . Finally, taking the supremum in (11) with respect to all sets $\Lambda \subset \Theta$ having probability at least $1 - \gamma$ yields the lower bound (8)

$$n^{-1} \log M_{\theta, \hat{\theta}} \geq \mathcal{C}(\gamma, \psi_{\theta|\hat{\theta}}, \hat{\theta}, P) - 2\delta \geq R - 2\delta, \quad (12)$$

for all rate $R \leq \mathcal{C}(\gamma, \psi_{\theta|\hat{\theta}}, \hat{\theta}, P)$ and $\theta \in \Lambda^*$, which is attained by some code with $\Lambda_\epsilon = \Lambda^*$. Next we prove the upper bound (9). From the converse part of Theorem 3.1, we have

$$n^{-1} \log M_{\theta, \hat{\theta}} \leq n^{-1} \log g_{\Lambda_\epsilon}(\mathcal{A}^n, \epsilon + \delta) - H(\mathcal{W}_{\Lambda_\epsilon}|P) + \delta, \quad (13)$$

for all $\theta \in \Lambda_\epsilon$. Since $\mathcal{A}^n = \mathcal{T}_{P|\hat{\theta}_T}^n(\delta, \hat{\theta})$ implies that any common $(\epsilon + \delta)$ -image of \mathcal{A}^n will be included in $\bigcap_{\theta \in \Lambda_\epsilon} \mathcal{T}_{W_\theta P}^n(\delta'_n)$, Lemma 1.2.12 in [5] ensures that there exists $n \geq n''_0(|\mathcal{X}|, |\mathcal{Y}|, \epsilon, \delta)$ such that,

$$n^{-1} \log g_{\Lambda_\epsilon}(\mathcal{A}^n, \epsilon + \delta) \leq \inf_{\theta \in \Lambda_\epsilon} H(W_\theta P) + \delta. \quad (14)$$

Then by applying equation (14) to equation (13), and then by taking its supremum with respect to all sets $\Lambda \subset \Theta$ having probability at least $1 - \gamma$, we obtain

$$n^{-1} \log M_{\theta, \hat{\theta}} \leq \mathcal{C}(\gamma, \psi_{\theta|\hat{\theta}}, \hat{\theta}, P) + 2\delta \leq R + 2\delta, \quad (15)$$

for all $R \geq \mathcal{C}(\gamma, \psi_{\theta|\hat{\theta}}, \hat{\theta}, P)$ and $\theta \in \Lambda_\epsilon$ with $\Pr(\theta \notin \Lambda_\epsilon|\hat{\theta}) < \gamma$, and this concludes the proof.

4. Numerical Results and Discussion

In this section, we illustrate our results via a realistic single user mobile wireless communication system involving a Ricean block flat fading channel, where the channel state is described by a single fading coefficient. The channel states in each block are i.i.d. and unknown at the transmitter and the receiver. The transmission extends over many blocks (coherence intervals) such that the average outage capacity (1) is indeed the appropriate performance criterion. The practical significance of this capacity stems from QoS requirements present in many communication services.

Within each block, the actual codeword (data) is preceded by a length- N training sequence $\mathbf{x}_T = [x_0, \dots, x_{N-1}]$ of power P_T which is known by the receiver. This enables maximum likelihood (ML) channel estimation of the fading coefficient θ at the receiver yielding the estimate $\hat{\theta}_R$. In many wireless systems, CSI at the transmitter $\hat{\theta}_T$ has to be provided by the receiver via a feedback/CSIT. This allows the transmitter to perform power

control $\mathcal{P}(\hat{\theta}_R)$, i.e., allocate more transmit power when the estimated channel is good, and less or no power when the channel is bad. Below, we consider the following three feedback schemes: (i) no feedback, i.e., absence of CSIT; (ii) an instantaneous and unlimited feedback/CSIT ($\hat{\theta}_T = \hat{\theta}_R$); (iii) an instantaneous and rate-limited feedback/CSIT; here the CSI is quantized using a quantization codebook which is known at the transmitter and the receiver (we construct this codebook using the well-known Lloyd-Max algorithm).

4.1. Channel Model

The channel model within a block is given by (all quantities are complex-valued) $Y[i] = H[i] X[i] + Z[i]$, where $X[i]$ and $Y[i]$ are the discrete-time transmit and receive signal, respectively, H is the fading coefficient, and $Z[i] \sim \mathcal{CN}(0, \sigma_Z^2)$ is i.i.d. zero-mean, circularly complex Gaussian noise. The transmit signal is subject to the average power constraint $E\{|X[i]|^2\} \leq \mathcal{P}(\hat{\theta}_T)$ with $E_{\hat{\theta}_T}\{\mathcal{P}(\hat{\theta}_T)\} \leq P$. The optimum power allocation $\mathcal{P}(\hat{\theta}_R)$ is obtained using Lagrange multipliers and the Kuhn-Tucker theorem. The channel state $\theta = H[i]$ is assumed to be circularly complex Gaussian $\theta \sim \psi(\theta) = \mathcal{CN}(\mu_h, 2\sigma_h^2)$. The Rice factor is defined as $K_h = \frac{|\mu_h|^2}{2\sigma_h^2}$. The ML estimate $\hat{\theta}_R = \hat{H}[i]$ is obtained by correlating the received signal with the known training sequence \mathbf{x}_T . Its performance can be characterized via the pdf $\psi(\theta|\hat{\theta}_R) = \mathcal{CN}(\rho\hat{\theta}_R + (1-\rho)\mu_h, \rho\sigma_W^2)$, where $\rho = \frac{2\sigma_h^2}{\sigma_W^2 + 2\sigma_h^2}$ and $\sigma_W^2 = \sigma_Z^2/(NP_T)$.

For a given estimate $\hat{\theta}_0$, to evaluate (3) requires solving an optimization problem where we have to determine the optimum set Λ^* , and the associated channel state $\theta^* \in \Lambda^*$ minimizing mutual information. The estimation-induced outage capacity can then be shown to be given by

$$C(\gamma, \psi_{\theta|\hat{\theta}}, \hat{\theta}_0) = \log_2 \left(1 + \frac{(r^*(\gamma, \hat{\theta}_0))^2 \mathcal{P}(\hat{\theta}_{T,0})}{\sigma_Z^2} \right), \text{ where } r^*$$

is the γ -percentile¹ of $\psi(r|\hat{\theta} = \hat{\theta}_0)$ with $r = |\theta|$ (for further details see [6]).

4.2. Results and Discussion

Fig. 1 shows the average estimation-induced outage capacity (cf. (1)) in bits per channel use for outage probability $\gamma = 0.01$ versus the signal-to-noise ratio $\text{SNR} = |\mu_h|^2 P / \sigma_Z^2$ for different amounts of training and for unlimited and absent feedback/CSIT (all numerical results were obtained using Monte Carlo simulations). For comparison, we show ergodic capacity under perfect CSI. The channel's Rice factor was $K_h = 0$ dB. It is seen that the average rate increases with the amount of CSIR and CSIT. To achieve

¹It can be computed by using the cumulative distribution of a non-central chi-square of two degrees of freedom.

1.5 bit per channel use without feedback/CSIT, it is seen that a scheme with estimated CSIR and $N = 3$ (∇ markers) requires 5 dB, i.e., 4.3 dB more than with perfect CSIR (solid line). Whereas if the training length is further reduced to $N = 1$ (\circ markers), this gap increases to 6.4 dB. In the case of unlimited feedback (CSIT=CSIT), the SNR requirements for 1.5 bit per channel use are -1.3 dB (perfect CSIR, dashed line), 2.1 dB (estimated CSIR with $N = 3$, \times markers), and 3.7 dB (estimated CSIR with $N = 1$, \times markers), respectively. Thus, with unlimited feedback the gap between estimated and perfect CSI is slightly smaller than without feedback (3.4 dB and 5 dB with $N = 3$ and $N = 1$, respectively). Observe that for values of SNR larger than 10 dB similar performance are achieved without feedback/CSIT and $N = 3$ comparing to a system with unlimited feedback and $N = 1$. Therefore, using this information a system designer may decide to use training sequences of length $N = 3$ instead of implementing a feedback channel.

Fig. 2 shows the average estimation-induced outage capacity for an outage probability $\gamma = 0.01$ and rate-limited feedback/CSIT versus the SNR. We suppose two bits of feedback ($R_{FB} = 2$), and training sequences of length $N = 3$. Observe that at 1.5 bits the gap between the average outage capacity without feedback and rate-limited feedback is 1 dB for two bits of feedback/CSIT. Whereas the gap respect to the average outage capacity with unlimited feedback is only 2 dB.

5. Conclusions

In this paper we have studied the problem of reliable communications over unknown DMCs when the receiver and transmitter only know an estimate of the channel state. We proposed to characterize the information theoretic limits of such scenarios in terms of the novel notion of estimation-induced outage capacity. We provided an explicit expression for the maximum achievable outage rate in the context of an associated coding theorem and its strong converse. We used a Ricean fading channel and maximum likelihood channel estimation to illustrate our approach by computing its mean outage capacity. Our results are useful to assess the amount of training data and feedback required to achieve a target rate satisfying a quality of service constraint. It will be attractive to study coding schemes achieving this capacity because this allows to design communication systems with QoS constraints and imperfect channel estimation.

Acknowledgment

The authors are grateful to Prof. Te Sun Han for many helpful discussions and suggestions on the technical aspects of the paper's proofs.

A. Auxiliary results

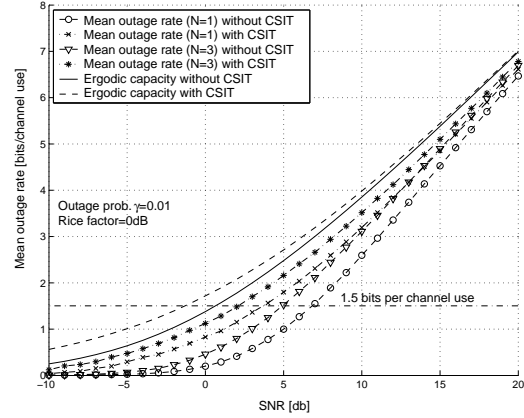


Figure 1: Average estimation-induced outage capacity for different amounts of training vs. SNR.

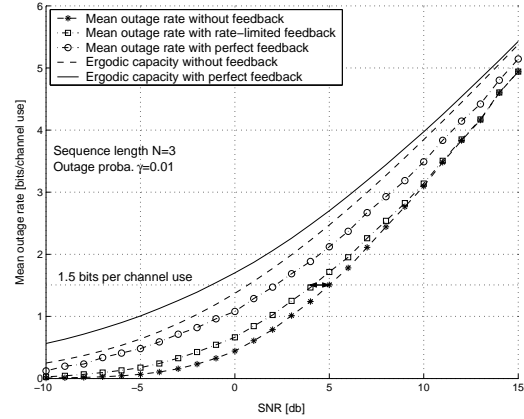


Figure 2: Average estimation-induced outage capacity with two bits of rate-limited feedback ($R_{FB} = 2$) vs. SNR.

This appendix introduces a few concepts and furthermore provides some auxiliary technical results required for the proof of Theorem 2.1.

Robust Decoders: Let $\mathcal{A}^n \subset \mathcal{X}^n$ denote a set of transmit sequences and let $W_\theta(\cdot|x) = W(\cdot|x, \theta)$. A set $\mathcal{B}^n \subset \mathcal{Y}^n$ (depending on $\Lambda \subset \Theta$) is called a *robust ϵ -decoding set* for a sequence $\mathbf{x} \in \mathcal{A}^n$ and an unknown DMC $W(\cdot|x, \theta) \in \mathcal{W}_\Theta$, if $\Pr(W^n(\mathcal{B}^n|\mathbf{x}, \theta) > 1 - \epsilon|\hat{\theta}) \geq 1 - \gamma$.

A set $\mathcal{B}^n \subset \mathcal{Y}^n$ of receive sequences is called a *common η -image* ($0 < \eta \leq 1$) of a transmit set $\mathcal{A}^n \subset \mathcal{X}^n$ for the collection of DMCs \mathcal{W}_Λ , iff $\inf_{\theta \in \Lambda} W^n(\mathcal{B}^n|\mathbf{x}, \theta) \geq \eta$ for all $\mathbf{x} \in \mathcal{A}^n$. Finally, $\Lambda \subset \Theta$ is called a *confidence set* for θ given $\hat{\theta}$, if $\Pr(\theta \notin \Lambda|\hat{\theta}) < \gamma$ and γ is the outage probability.

Robust I-Typical Sets: A robust I-typical set is defined as $\mathcal{B}_\Lambda^n(\mathbf{x}, \delta_n) = \bigcup_{\theta \in \Lambda} \mathcal{T}_{W_\theta}^n(\mathbf{x}, \delta_n)$, with arbitrary $\Lambda \subset \Theta$ and δ -sequence $\{\delta_n\}$ (cf. [5]).

Lemma A.1 For any $0 < \gamma, \epsilon < 1$, a necessary and sufficient condition for a robust I-typical set $\mathcal{B}_\Lambda^n(\mathbf{x}, \delta_n)$ to be a robust ϵ -decoding set with probability $1 - \gamma$ is that Λ be a confidence set.

Theorem A.2 For any collection of DMCs \mathcal{W}_Λ and associated robust I-typical set $\mathcal{B}_\Lambda^n(\mathbf{x}, \delta_n)$ with $\mathbf{x} \in \mathcal{T}_P^n(\mathbf{x}, \delta_n)$, there exists an index n_0 such that for all $n \geq n_0$ the size $\|\mathcal{B}_\Lambda^n(\mathbf{x}, \delta_n)\|$ of the robust I-typical set is bounded as

$$\left| \frac{1}{n} \log \|\mathcal{B}_\Lambda^n(\mathbf{x}, \delta_n)\| - H(\mathcal{W}_\Lambda|P) \right| \leq \eta_n.$$

Here, $H(\mathcal{W}_\Lambda|P) = \sup_{V \in \mathcal{W}_\Lambda} H(V|P)$ and $\eta_n \rightarrow 0$ as $\delta_n \rightarrow 0$ and $n \rightarrow \infty$.

Proof: We first show that the size of $\mathcal{B}_\Lambda^n(\mathbf{x}, \delta_n)$ is asymptotically equal to the size of $\mathcal{B}_\Sigma^n(\mathbf{x}) = \bigcup_{V \in \Sigma} \mathcal{T}_V^n(\mathbf{x})$ where $\Sigma = \mathcal{W}_\Lambda \cap \mathcal{P}_n(\mathcal{Y})$ is the intersection of \mathcal{W}_Λ with the set $\mathcal{P}_n(\mathcal{Y})$ of empirical distributions induced by receive sequences of length n . In particular, there exists an index n_0 such that for all $n \geq n_0$ and $\mathbf{x} \in \mathcal{X}^n$

$$\|\mathcal{B}_\Sigma^n(\mathbf{x})\| \leq \|\mathcal{B}_\Lambda^n(\mathbf{x}, \delta_n)\| \leq (1+n)^{|\mathcal{X}||\mathcal{Y}|} \|\mathcal{B}_\Sigma^n(\mathbf{x})\|. \quad (16)$$

The lower bound in (16) is trivial. We will next establish that there exists $\epsilon_n > 0$ such that for all $n \geq n_0$

$$\bigcup_{W \in \mathcal{W}_\Lambda} \mathcal{T}_W^n(\mathbf{x}, \delta_n) \subseteq \bigcup_{V \in \Sigma} \mathcal{T}_V^n(\mathbf{x}, \epsilon_n), \quad (17)$$

from which the upper bound in (16) follows via basic properties of types (cf. [5]).

Assume that \mathcal{W}_Λ is a relatively τ_0 -open subset of $\mathcal{W}_\Lambda \cup \mathcal{P}_n(\mathcal{Y})$, i.e., every $W \in \mathcal{W}_\Lambda$ has a τ_0 -neighborhood defined in the τ_0 -topology [7]. Then there exists n_0 such that for any $n \geq n_0$ and $\epsilon > 0$, the ϵ -open ball $U_0(W, \epsilon)$ satisfies $U_0(W, \epsilon) \cap \mathcal{P}_n(\mathcal{Y}) \subset \mathcal{W}_\Lambda$. Choose $0 < \epsilon' < \epsilon$ and pick an empirical conditional PM $V \in \mathcal{P}_n(\mathcal{Y})$ such that for all $(a, b) \in \mathcal{X} \times \mathcal{Y}$, $|V(b|a) - W(b|a)| < \epsilon'_n$ and $V(b|a) = 0$ if $W(b|a) = 0$. The continuity properties of information divergences imply that for any sequence $\mathbf{y} \in \mathcal{T}_W^n(\mathbf{x}, \delta_n)$ (i.e., $\mathcal{D}(\hat{W}_n \| W | \hat{P}_n) \leq \delta_n$), $|\hat{W}_n(b|a)\hat{P}_n(a) - W(b|a)\hat{P}_n(a)| \leq \sqrt{\delta_n}/2$ and hence $|\hat{W}_n(b|a)\hat{P}_n(a) - V(b|a)\hat{P}_n(a)| \leq \epsilon' + \sqrt{\delta_n}/2$. Finally, from this equation it is easy to show that there exists an $\epsilon_n > 0$ such that $\mathcal{D}(\hat{W}_n \| V | \hat{P}_n) \leq \epsilon_n$, i.e., $\mathbf{y} \in \mathcal{T}_V^n(\mathbf{x}, \epsilon_n)$. Consequently, for any $W \in \mathcal{W}_\Lambda$ and large enough n , it is possible to find $V \in \Sigma$ and $\epsilon_n > 0$ such that $\mathcal{T}_W^n(\mathbf{x}, \delta_n) \subseteq \mathcal{T}_V^n(\mathbf{x}, \epsilon_n)$, thus establishing (17). Using similar arguments as above and the uniform continuity of the entropy function, it can be shown that there exists n'_0 such that for all $n \geq n'_0$ and $\mathbf{x} \in \mathcal{X}^n$

$$\left| \frac{1}{n} \log \|\mathcal{B}_\Sigma^n(\mathbf{x})\| - \sup_{V \in \mathcal{W}_\Lambda} H(V|P) \right| \leq \xi_n, \quad (18)$$

with $\xi_n = |\mathcal{X}||\mathcal{Y}|n^{-1} \log(n+1) + \xi'_n$ and $\xi'_n \rightarrow 0$ as $n \rightarrow \infty$. The theorem follows by combining the inequalities (16) and (18) and setting $\eta_n = \xi_n + |\mathcal{X}||\mathcal{Y}|n^{-1} \log(n+1)$.

Proof of Theorem 3.1: To prove the direct part, consider an admissible code that is maximal, i.e., it cannot be extended by arbitrary $(\mathbf{x}_{M+1}; \mathcal{D}_{M+1}^n)$ such that the extended code remains admissible. Define the set $\mathcal{D}^n = \bigcup_{i=1}^M \mathcal{D}_i^n$

with $\mathcal{D}_i^n \subseteq \mathcal{B}_\Lambda^n(\mathbf{x}_i, \delta)$, and choose $\delta < \epsilon$ such that $1 - \epsilon > \epsilon - \delta$. Then $\inf_{\theta \in \Lambda} W^n(\mathcal{D}^n | \mathbf{x}_i, \theta) > \epsilon - \delta$, (19)

for all $\mathbf{x}_i \in \mathcal{X}^n$. As the code is maximal, for all $\mathbf{x} \in \mathcal{X}^n \setminus \{\mathbf{x}_1, \dots, \mathbf{x}_M\}$, we have $\inf_{\theta \in \Lambda} W^n(\mathcal{B}_\Lambda^n \setminus \mathcal{D}^n | \mathbf{x}, \theta) \leq 1 - \epsilon$.

This equation implies that for all $\theta \in \Lambda$ and large enough n

$$W^n(\mathcal{D}^n | \mathbf{x}, \theta) \geq \epsilon - \delta, \quad (20)$$

for all $\mathbf{x} \in \mathcal{X}^n \setminus \{\mathbf{x}_1, \dots, \mathbf{x}_M\}$. The inequalities (19) and (20) together imply that \mathcal{D}^n is a common $(\epsilon - \delta)$ -image of the set \mathcal{X}^n via the collection of channels \mathcal{W}_Λ . By the definition of $\mathbf{g}_\Lambda(\mathcal{X}^n, \epsilon - \delta)$ it follows that

$$\|\mathcal{D}^n\| \geq \mathbf{g}_\Lambda(\mathcal{X}^n, \epsilon - \delta). \quad (21)$$

On the other hand, $\mathcal{D}_i^n \subseteq \mathcal{B}_\Lambda^n(\mathbf{x}_i, \delta)$ implies that

$$\|\mathcal{D}^n\| \leq M_{\theta, \delta} \exp[n(H(\mathcal{W}_\Lambda|P) + \delta)], \quad (22)$$

for n large enough and all $\theta \in \Lambda$, where the last inequality follows by applying the cardinality upper bound of Theorem A.2. The lower bound (6) is then immediately obtained by combining (21) and (22). To prove the converse part, let $\hat{\mathcal{D}}^n$ be a common $(\epsilon + \delta)$ -image via the collection of channels $\mathcal{W}_{\Lambda_\epsilon}$, i.e.,

$$\inf_{\theta \in \Lambda_\epsilon} W^n(\hat{\mathcal{D}}^n | \mathbf{x}_m, \theta) \geq \epsilon + \delta, \quad \text{for } m \in \mathcal{M}, \quad (23)$$

that achieves the minimum in (4), i.e., $\|\hat{\mathcal{D}}^n\| = \mathbf{g}_{\Lambda_\epsilon}(\mathcal{X}^n, \epsilon + \delta)$. For any admissible code, (5) and (23) imply $\inf_{\theta \in \Lambda_\epsilon} W^n(\mathcal{D}_m^n \cap \hat{\mathcal{D}}^n | \mathbf{x}_m, \theta) \geq \delta$ for $m \in \mathcal{M}$. Using Corollary 1.2.14 in [5], we hence obtain

$$\|\mathcal{D}_m^n \cap \hat{\mathcal{D}}^n\| \geq \exp[n(H(\mathcal{W}_{\Lambda_\epsilon}|P) - \delta)], \quad (24)$$

for n large enough. On the other hand, the decoding sets \mathcal{D}_m^n are disjoint and thus

$\mathbf{g}_{\Lambda_\epsilon}(\mathcal{X}^n, \epsilon + \delta) = \|\hat{\mathcal{D}}^n\| \geq M_{\theta, \delta} \exp[n(H(\mathcal{W}_{\Lambda_\epsilon}|P) - \delta)]$, where the last inequality follows from (24). This inequality is equivalent to (7) and concludes the proof of the theorem.

References

- [1] A. Lapidoth, "Reliable communication under channel uncertainty," *IEEE Trans. Inform. Theory*, vol. IT-44, pp. 2148–2177, October 1998.
- [2] M. Médard, "The effect upon channel capacity in wireless communication of perfect and imperfect knowledge of the channel," *IEEE Trans. Inform. Theory*, vol. IT-46, no. 3, pp. 933–946, May 2000.
- [3] L. Ozarow, S. Shamai, and A. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Inform. Theory*, vol. 43, no. 2, pp. 359–378, May 1994.
- [4] A. Goldsmith and P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. Inform. Theory*, vol. IT-43, pp. 1986–1992, 1997.
- [5] I. Csiszár and J. Körner, *Information theory: coding theorems for discrete memoryless systems*. Academic, New York, 1981.
- [6] P. Piantanida, G. Matz, and P. Duhamel, "Estimation-induced outage capacity of ricean channels," in *Proc. of Signal Processing for Advanced Wireless Communications (SPAWC)*, July 2006.
- [7] I. Csiszár, "Sanov property, generalize I-projection and a conditional limit theorem," *Ann. Probability*, vol. 12, pp. 768–793, 1984.